


RESEARCH ARTICLE

Open Access

# Limited overlap in significant hits between genome-wide association studies on two airflow obstruction definitions in the same population



Diana A. van der Plaat<sup>1,2</sup>, Judith M. Vonk<sup>1,2</sup>, Lies Lahousse<sup>3,4</sup>, Kim de Jong<sup>1,2</sup>, Alen Faiz<sup>5,2</sup>, Ivana Nedeljkovic<sup>3</sup>, Najaf Amin<sup>3</sup>, Cleo C. van Diemen<sup>6</sup>, Guy G. Brusselle<sup>3,4,7</sup>, Yohan Bossé<sup>8</sup>, Corry-Anke Brandsma<sup>5,2</sup>, Ke Hao<sup>9</sup>, Peter D. Paré<sup>10</sup>, Cornelia M. van Duijn<sup>3</sup>, Dirkje S. Postma<sup>11,2</sup> and H. Marika Boezen<sup>1,2\*</sup> 

## Abstract

**Background:** Airflow obstruction is a hallmark of chronic obstructive pulmonary disease (COPD), and is defined as either the ratio between forced expiratory volume in one second and forced vital capacity (FEV<sub>1</sub>/FVC) < 70% or < lower limit of normal (LLN). This study aimed to assess the overlap between genome-wide association studies (GWAS) on airflow obstruction using these two definitions in the same population stratified by smoking.

**Methods:** GWASes were performed in the LifeLines Cohort Study for both airflow obstruction definitions in never-smokers (NS = 5071) and ever-smokers (ES = 4855). The FEV<sub>1</sub>/FVC < 70% models were adjusted for sex, age, and height; FEV<sub>1</sub>/FVC < LLN models were not adjusted. Ever-smokers models were additionally adjusted for pack-years and current-smoking. The overlap in significantly associated SNPs between the two definitions and never/ever-smokers was assessed using several *p*-value thresholds. To quantify the agreement, the Pearson correlation coefficient was calculated between the *p*-values and ORs. Replication was performed in the Vlagtwedde-Vlaardingen study (NS = 432, ES = 823). The overlapping SNPs with *p* < 10<sup>-4</sup> were validated in the Vlagtwedde-Vlaardingen and Rotterdam Study cohorts (NS = 1966, ES = 3134) and analysed for expression quantitative trait loci (eQTL) in lung tissue (*n* = 1087).

**Results:** In the LifeLines cohort, 96% and 93% of the never- and ever-smokers were classified concordantly based on the two definitions. 26 and 29% of the investigated SNPs were overlapping at *p* < 0.05 in never- and ever-smokers, respectively. At *p* < 10<sup>-4</sup> the overlap was 4% and 6% respectively, which could be change findings as shown by simulation studies. The effect estimates of the SNPs of the two definitions correlated strongly, but the *p*-values showed more variation and correlated only moderately. Similar observations were made in the Vlagtwedde-Vlaardingen study. Two overlapping SNPs in never-smokers (*NFYC* and *FABP7*) had the same direction of effect in the validation cohorts and the *NFYC* SNP was an eQTL for *NFYC-AS1*. *NFYC* is a transcription factor that binds to several known COPD genes, and *FABP7* may be involved in abnormal pulmonary development.

(Continued on next page)

\* Correspondence: [h.m.boezen@umcg.nl](mailto:h.m.boezen@umcg.nl)

<sup>1</sup>Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9700, RB, Groningen, The Netherlands

<sup>2</sup>Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands

Full list of author information is available at the end of the article



(Continued from previous page)

**Conclusions:** The definition of airflow obstruction and the population under study may be important determinants of which SNPs are associated with airflow obstruction. The genes *FABP7* and *NFYC(-AS1)* could play a role in airflow obstruction in never-smokers specifically.

**Keywords:** Genome-wide association study, Genetics, Airflow obstruction, COPD

## Background

Chronic obstructive pulmonary disease (COPD) is a major cause of morbidity and mortality in the world and encompasses emphysema, chronic bronchitis, and small airways disease [1, 2]. The diagnosis of COPD is largely based on the presence of airflow obstruction, measured by the spirometric assessment (post-bronchodilator) of the ratio between forced expiratory volume in one second and forced vital capacity ( $FEV_1/FVC$ ). The Global initiative for chronic Obstructive Lung Disease (GOLD) recommends to use a fixed cut-off for defining airflow obstruction, namely an  $FEV_1/FVC$  ratio below 70% [3], whereas the American Thoracic Society/European Respiratory Society (ATS/ERS) guidelines recommend to define airflow obstruction as  $FEV_1/FVC$  below the lower limit of normal (LLN) [4]. The LLN is a reference value based on sex, age, height and ethnicity and is calculated as the lower fifth percentile of a healthy reference population [5]. There is a considerable controversy about which definition should be used in research and clinical practice, since both may lead to misclassifications [5–8]. This has important implications, since misclassifications may lead to inappropriate medication and therapies [9, 10].

It is generally accepted that both genetic susceptibility and environmental factors contribute to airflow obstruction. Genetic variants associated with airflow obstruction have been identified by several genome-wide association studies (GWAS), but different definitions of airflow obstruction and populations were used. [11–16] As an illustration, the case-control study including only smokers with >2.5 pack-years by Pillai et al. used the fixed ratio ( $FEV_1/FVC < 70%$ ) to define airflow obstruction, while the population based study including both ever- and never-smokers by Wilk et al. used the lower limit of normal (LLN) [15, 16]. Only few regions were identified in both studies, namely the *CHRNA5/3* and *HHIP* regions. We therefore aimed to assess the genetic overlap between the two definitions of airflow obstruction in the same individuals. We stratified by smoking status to assess the overlap between the two airflow obstruction definitions in never- and ever-smokers separately. We used the Lifelines Cohort Study as discovery sample and the Vlagtwedde-Vlaardingen study to replicate our observations. In addition, genetic loci associated with both airflow obstruction definitions could indicate robust genetic associations with airflow obstruction,

which could potentially be novel loci. We therefore, as a secondary aim, validated the top overlapping single-nucleotide polymorphisms (SNPs) between the two airflow obstruction discovery analyses in an independent SNP validation sample and assessed if they were acting as expression quantitative trait loci (eQTLs) in a lung tissue sample.

## Materials and methods

### Study populations

To study the overlap between the two airflow obstruction definitions, all subjects with available genotypic data were included from the Dutch LifeLines Cohort Study (discovery sample) and the Vlagtwedde-Vlaardingen study (replication sample) [17–19]. In addition, subjects from the Vlagtwedde-Vlaardingen study and the three independent cohorts of the Rotterdam Study (RS I to III) were selected to validate the top overlapping SNPs from LifeLines (SNP validation sample), thereby increasing the SNP validation sample size [20]. All subjects provided written informed consent and the studies were approved by local medical ethics committees. Smoking status was based on self-reported smoking history and pack-years smoked. In the stratified analyses never-smokers having smoked 0 pack-years and ever-smokers having smoked >5 pack-years were included, thereby excluding subjects with >0 and  $\leq 5$  pack-years. Subjects were defined as having airflow obstruction based on having a pre-bronchodilator  $FEV_1/FVC$  ratio (%) <70% or <LLN (based on Global Lung Initiative 2012 (GLI-2012)) [21]. All subjects completed pulmonary function testing according to ATS or ERS criteria [22]. Additional details are provided in Additional file 1.

### Genotyping

The IlluminaCytoSNP-12 arrays were used to genotype blood samples in LifeLines and the Vlagtwedde-Vlaardingen study. SNPs with a genotype call-rate  $\geq 95%$ , minor allele frequency  $\geq 1%$  and Hardy-Weinberg  $p$ -value  $\geq 10^{-4}$  were included. Non-Caucasian samples and first-degree relatives were excluded based on self-reporting, outlier (Identity By State) and principal component analysis. After quality control, 227,981 genotyped SNPs were included in the discovery analyses (LifeLines) and 242,926 genotyped SNPs were included

in the replication analyses (Vlagtwedde-Vlaardingen). Only genotyped SNPs were included in the analyses to prevent introducing bias, since it is known that imputation can reduce the effect size estimation, especially if healthy controls are used as reference [23]. Blood samples in the Rotterdam study were genotyped with the 610K and 660K Illumina arrays and similar QC criteria as in the other cohorts were applied.

### Statistical analysis

Four separate GWASes were performed assessing the genetic associations between the two definitions of airflow obstruction, stratified by smoking status, for both the LifeLines (discovery) and Vlagtwedde-Vlaardingen (replication) studies. Logistic regression (additive genetic model) was performed using PLINK (v1.07) [24]. The “FEV<sub>1</sub>/FVC < 70%” model was adjusted for sex, age and height. The “FEV<sub>1</sub>/FVC < LLN” model was not adjusted for these variables, since they are included in the LLN calculation. In ever-smokers, the models were additionally adjusted for pack-years and current-smoking. We used different *p*-value thresholds to assess the number of overlapping SNPs between the two definitions. In addition, to quantify the agreement of the results between the two definitions and between never- and ever-smokers, we calculated the Pearson correlation coefficient between the *p*-values and between the ORs.

### Power simulations

Our study has a relative small sample size ( $n = 5070$ ) and therefore relative low power. We assessed the effect of low power on the overlap between the two definitions by increasing our never-smoking discovery sample (LifeLines) 2 ( $n = 10,140$ ) and 4 ( $n = 20,280$ ) times. In addition, to assess if our results were spurious, we used our never-smoking discovery sample and randomly allocated 10 times the airflow obstruction cases but keeping the same distribution as in our original dataset (FEV<sub>1</sub>/FVC < 70%:  $n = 548$ , FEV<sub>1</sub>/FVC < LLN:  $n = 401$ , overlapping cases:  $n = 371$  (64%)). For both simulation studies, we repeated the GWAS analyses on both airflow obstruction definitions in the created datasets and compared the number of overlapping SNPs.

### Validation of overlapping SNPs

Only the top overlapping SNPs between the two airflow obstruction definitions in the discovery sample (LifeLines) were evaluated in the SNP validation sample, the Vlagtwedde-Vlaardingen study and RS I to III. A fixed-effects meta-analysis of the effect estimates weighted by the inverse of the standard errors from all four validation cohorts was performed using METAL (v2011) [25]. We considered replication if the meta-analysis *p*-value was below the Bonferroni corrected *p*-value defined as

0.05/number of overlapping SNPs and, in addition, had the same direction of effect in all cohorts. In addition, SNP\*ever-smoking interactions were estimated and we assessed if the overlapping SNPs were associated with gene expression levels in lung tissue within a 4 Mb window around the SNP (2 Mb on either side of the SNP), using data from the lung eQTL consortium [26]. In total, 1087 subjects were included in the linear regression model, adjusted for disease status, age, sex, smoking, and cohort specific principal components. SNPs with a *p*-value below the Bonferroni corrected threshold ( $p = 0.05/\text{number of probesets}$ ) were considered significant eQTLs. See Additional file 1 and GEO accession numbers GSE23546 and GPL10379 for additional information.

## Results

### Population characteristics

The LifeLines cohort (discovery sample) included 5070 never-smokers and 4855 ever-smokers with complete data on all covariates (see Table 1). Of the never-smokers in LifeLines, 96% had a concordant airflow obstruction classification for the two definitions: 89% did not have airflow obstruction and 7% did have airflow obstruction. The remaining 4% had a discordant classification (see Additional file 1: Table S1A). Figure 1a shows that of all never-smoking subjects with airflow obstruction based on at least one airflow obstruction definition ( $n = 578$ ), 36% had a discordant airflow obstruction classification. Of the ever-smokers, 93% was classified concordantly: 77% did not have airflow obstruction and 17% did have airflow obstruction. The remaining 7% had a discordant classification (see Additional file 1: Table S1B). Of all ever-smoking subjects with airflow obstruction based on at least one definition ( $n = 1138$ ), 30% had a discordant airflow obstruction classification (see Fig. 1a). Subjects with an FEV<sub>1</sub>/FVC < 70% and > LLN were aged between 41 and 85, and subjects with an FEV<sub>1</sub>/FVC > 70% and < LLN were aged between 22 and 43. These and other characteristics of the airflow obstruction groups separately for never- and ever-smokers in LifeLines are shown in Additional file 1: Table S2.

The Vlagtwedde-Vlaardingen study (replication sample) included 432 never-smokers and 823 ever-smokers (see Table 1). Of the Vlagtwedde-Vlaardingen study, 94% and 90% of the never- and ever-smokers were classified concordantly based on the two definitions (see Additional file 1: Table S1 C-D).

The SNP validation sample used for the SNP validation meta-analysis included 1966 never-smokers and 3134 ever-smokers from the Vlagtwedde-Vlaardingen study and RS I to III (see Table 1).

**Table 1** Characteristics of never- and ever-smokers included in the current study

	Never-smokers					Ever-smokers				
	LifeLines	Vla-Vla <sup>b</sup>	RS I	RS II	RS III	LifeLines	Vla-Vla <sup>b</sup>	RS I	RS II	RS III
N with no missing data	5070	432	408	379	747	4855	823	640	583	1088
Males, N (%)	1942 (38)	103 (23)	81 (20)	124 (33)	308 (41)	2312 (48)	590 (72)	375 (59)	343 (59)	531 (49)
Age (yrs), median (min-max)	46 (18–89)	54 (36–79)	78 (72–94)	71 (65–98)	62 (51–93)	49 (22–85)	53 (35–79)	79 (72–95)	71 (65–93)	62 (52–93)
Height (cm), mean (SD)	174 (9)	166 (9)	163 (9)	166 (9)	171 (9)	175 (9)	173 (8)	168 (9)	171 (9)	172 (9)
Current-smokers, N (%)	–	–	–	–	–	2171 (45)	478 (58)	99 (16)	108 (19)	298 (27)
Pack-years (yrs), mean (SD)	–	–	–	–	–	17 (11)	27 (21)	29 (21)	29 (21)	27 (20)
Pulmonary function, mean (SD)										
FEV <sub>1</sub> %predicted (%), mean (SD) <sup>a</sup>	98 (13)	91 (13)	101 (18)	102 (16)	101 (16)	94 (14)	85 (15)	98 (23)	96 (21)	99 (19)
FEV <sub>1</sub> /FVC (%), mean (SD) <sup>b</sup>	78 (7)	76 (7)	77 (7)	79 (6)	78 (6)	75 (8)	72 (10)	74 (9)	74 (9)	75 (8)
FEV <sub>1</sub> /FVC <70%, N (%)	548 (11)	77 (18)	51 (13)	27 (7)	75 (10)	1107 (23)	287 (35)	163 (26)	151 (26)	235 (22)
FEV <sub>1</sub> /FVC < LLN, N (%)	401 (8)	49 (11)	9 (2)	9 (2)	25 (3)	833 (17)	212 (26)	63 (10)	62 (11)	122 (11)
Both FEV <sub>1</sub> /FVC <70% and < LLN, N (%)	371 (7)	49 (11)	9 (2)	9 (2)	25 (3)	802 (17)	207 (25)	63 (10)	62 (11)	122 (11)
Moderate/severe COPD, N (%) <sup>c</sup>	106 (2)	26 (6)	21 (5)	9 (2)	16 (2)	310 (6)	145 (18)	90 (14)	90 (15)	104 (10)

Discovery sample = LifeLines cohort study, replication samples = Vlagtwedde-Vlaardingen (Vla-Vla) study, and SNP validation sample = Vla-Vla and RS I to III

FEV<sub>1</sub> forced expiratory volume in one second, FVC forced vital capacity

<sup>a</sup>FEV<sub>1</sub>%predicted is based on the reference equation by GLI-2012 [21]

<sup>b</sup>FEV<sub>1</sub>%IVC (inspiratory vital capacity) for the Vlagtwedde-Vlaardingen study (Vla-Vla)

<sup>c</sup>COPD GOLD stage 2 and up (FEV<sub>1</sub>/FVC <70% and FEV<sub>1</sub>%p < 80% based on pre-bronchodilator measurements)

### GWAS results

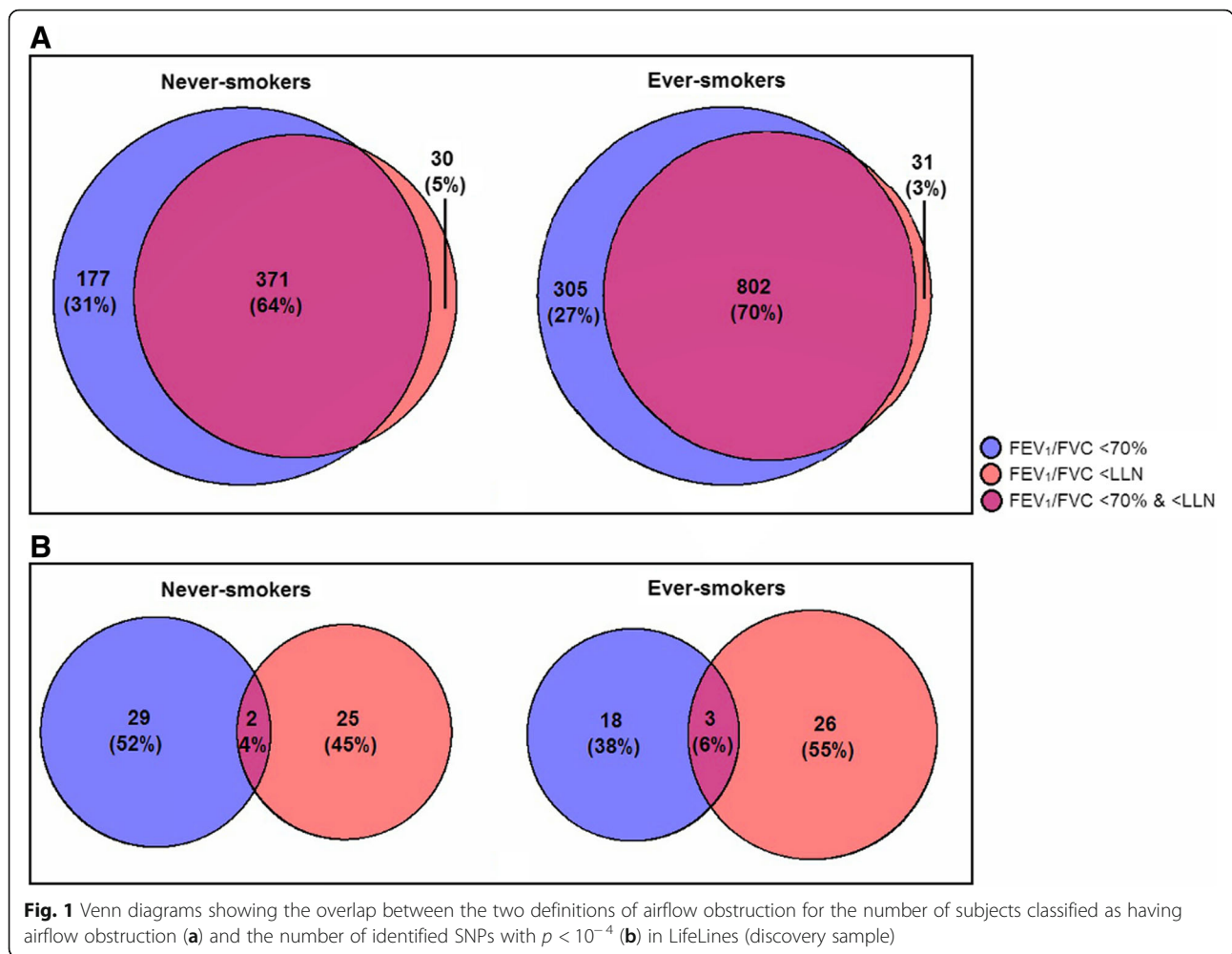
There was minimal population stratification in all analyses of LifeLines, indicated by the genomic inflation factor lambda ( $\lambda$ : 1.0002–1.0217, see Additional file 1: Figure S1). The results based on a  $p < 10^{-4}$  of all four analyses in LifeLines are given in See (Additional file 1: Tables S3-S6), including the Manhattan plots (see Additional file 1: Figures S2 and S3). For comparison, the effect estimates of both airflow obstruction definition are given in these tables. Summary statistics ( $p$  values, betas, and standard errors for all SNPs that were tested) of the GWAS result of both the Lifelines Cohort Study and the Vlagtwedde-Vlaardingen study are provided in Additional file 2.

### Overlap between the results

We used several  $p$ -value thresholds to assess the overlap between the GWAS results of both airflow obstruction definitions separately in never- and ever-smokers of LifeLines (see Table 2). A threshold of 0.05 resulted in the observation that 26% and 29% of the SNPs were overlapping between the two airflow obstruction definitions in never- and ever-smokers, respectively. Three percent of the SNPs were overlapping between never- and ever-smokers for both definitions. A smaller  $p$ -value threshold resulted in a lower percentage of overlap e.g. a threshold of  $p < 10^{-4}$  resulted in 4% and 6% overlapping

SNPs between the two airflow obstruction definitions in never- and ever-smokers, respectively (see Fig. 1b), and zero overlap between never- and ever-smokers using the same definition of airflow obstruction. Similar observations were made in the replication sample the Vlagtwedde-Vlaardingen study (see Table 2), since at  $p < 0.05$  the overlap between the definitions was 24% and 25% in never- and ever-smokers, respectively, and 2% of the SNPs were overlapping between never- and ever-smokers for both definitions.

The correlations between the SNP-specific  $p$ -values and ORs from the two airflow obstruction definitions were 0.48 ( $p$ -value) and 0.78 (OR) in never-smokers, and 0.51 ( $p$ -value) and 0.81 (OR) in ever-smokers (see Fig. 2). Between never- and ever-smokers the correlations of the SNP-specific  $p$ -values were 0.0008 for FEV<sub>1</sub>/FVC < 70% and 0.002 for FEV<sub>1</sub>/FVC < LLN, and for the OR the correlation was –0.02 for both definitions. Similar observations were made in the replication sample, the Vlagtwedde-Vlaardingen study. The correlations between the two definitions were 0.45 ( $p$ -value) and 0.76 (OR) in never-smokers, and 0.41 ( $p$ -value) and 0.74 (OR) in ever-smokers. Between never- and ever-smokers the correlations were –0.001 ( $p$ -value) and 0.015 (OR) for FEV<sub>1</sub>/FVC < 70% and –0.003 ( $p$ -value) and 0.004 (OR) for FEV<sub>1</sub>/FVC < LLN.



**Power simulations**

We found that the percentage of overlap increased when we expanded our never-smoking identification sample 2 and 4 times (see Additional file 1: Table S7). The overlap between SNPs with  $p < 10^{-4}$  was 3.6% in the original dataset, 16.5% in the 2x dataset and 26.6% in the 4x dataset. In addition, when we randomly allocated cases 10 times in our identification sample, we found the percentage of overlap between the two definitions at  $p < 10^{-4}$  varied between 0 to 16%, compared to 4% in the original dataset (see Additional file 1: Table S8).

**Validation of overlapping SNPs**

In never-smokers of LifeLines, two SNPs were overlapping between the FEV<sub>1</sub>/FVC < 70% and < LLN definitions at a threshold of  $p < 10^{-4}$  (see Table 3). The first SNP (rs7519348) is located in an intron of the gene nuclear transcription factor Y subunit C (*NFYC*), and the second SNP (rs6913003) is located in an intron of fatty acid binding protein 7 (*FABP7*, see Additional file 1: Figure S4 and S5 for LocusZoom plots). The minor alleles

of both SNPs were associated with a higher risk of airflow obstruction and had comparable odds ratios in both analyses. The SNP in *NFYC* (rs7519348) was significantly associated with FEV<sub>1</sub>/FVC < LLN in the SNP validation meta-analysis ( $p = 0.034$ ), but did not pass the multiple testing correction ( $0.05/2 = 0.025$ ), and was not significantly associated with FEV<sub>1</sub>/FVC < 70% in the SNP validation meta-analysis ( $p = 0.07$ ). The SNP in *FABP7* (rs6913003) was not significantly associated with FEV<sub>1</sub>/FVC < 70% or < LLN in the SNP validation meta-analyses ( $p = 0.08$  in both), although the direction of effect was the same in all independent cohorts. Both SNPs did not reach genome-wide significance according to the Bonferroni-corrected threshold ( $p < 2.19 \times 10^{-7}$ ) in the discovery analysis (LifeLines) or meta-analysis of both the discovery and SNP validation samples (see Table 3 and Additional file 1: Table S9). Yet, the odds ratios were comparable between all analyses (see Additional file 1: Table S10 and Figure S6). These two overlapping SNPs were not associated with airflow obstruction in ever-smokers and these associations were significantly different

**Table 2** Table showing the number of SNPs with a *p*-value below the mentioned threshold for both FEV<sub>1</sub>/FVC < 70% and < LLN analysis and the overlap

Threshold	Never-smokers			Ever-smokers (> 5 py)			Overlap never- and ever-smokers	
	70%	LLN	Overlap	70%	LLN	Overlap	70%	LLN
<i>Discovery analysis (LifeLines)</i>								
< 0.05	11,377	11,475	4755 (26%)	12,114	12,366	5445 (29%)	625 (2.7%)	616 (2.7%)
< 0.01	2232	2197	673 (18%)	2522	2493	824 (20%)	23 (0.5%)	21 (0.4%)
< 10 <sup>-3</sup>	222	233	54 (13%)	246	266	76 (17%)	1 (0.2%)	0
< 10 <sup>-4</sup>	31	27	2 (4%)	21	29	3 (6%)	0	0
< 10 <sup>-5</sup>	4	4	0	3	2	1 (25%)	0	0
< 10 <sup>-6</sup>	1	0	0	0	0	0	0	0
< Bonferroni*	0	0	0	0	0	0	0	0
<i>Replication analysis (Vlagtwedde-Vlaardingen study)</i>								
< 0.05	10,702	10,295	4026 (24%)	12,592	12,571	4976 (25%)	487 (2.1%)	488 (2.2%)
< 0.01	1857	1807	544 (17%)	2567	2609	754 (17%)	17 (0.4%)	12 (0.3%)
< 10 <sup>-3</sup>	164	174	38 (13%)	223	262	41 (9%)	0	0
< 10 <sup>-4</sup>	18	19	6 (19%)	12	21	0	0	0
< 10 <sup>-5</sup>	7	3	2 (25%)	0	0	0	0	0
< 10 <sup>-6</sup>	1	0	0	0	0	0	0	0
< Bonferroni*	0	0	0	0	0	0	0	0

\**P* < 2.19 × 10<sup>-7</sup>

between ever- and never-smokers as shown in the interaction analysis (see Additional file 1: Table S11).

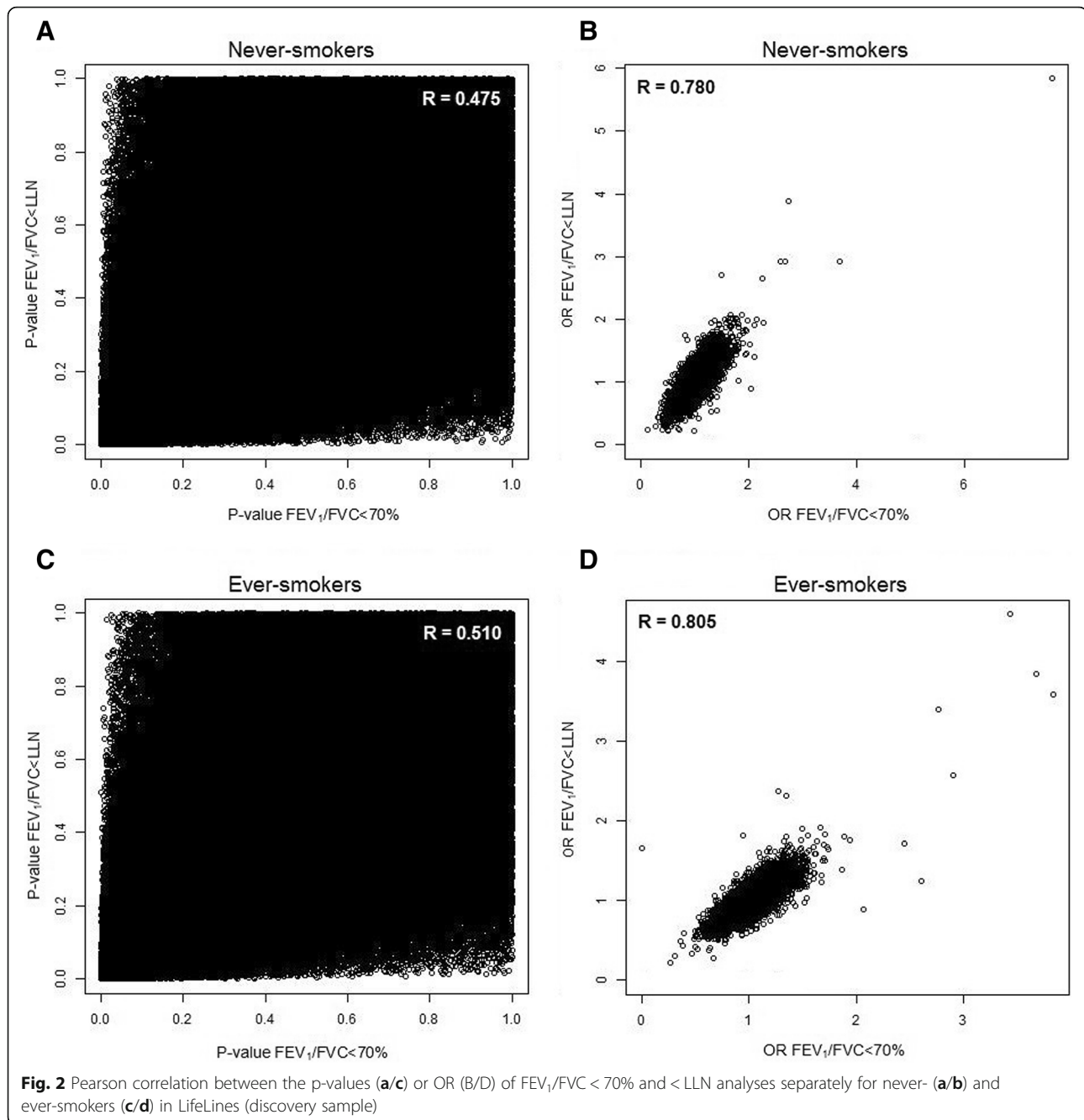
In ever-smokers of LifeLines, three SNPs were overlapping between the two analyses in at *p* < 10<sup>-4</sup> (see Table 3). The first SNP (rs13118083) is annotated to hedgehog interacting protein (*HHIP*, 342 kb away), but is located within the long non-coding RNA *LOC105377462* according to the SNP database by NCBI (<https://www.ncbi.nlm.nih.gov/SNP/>). The second SNP (rs7074210) is located approximately 62 kb from ST8 Alpha-N-Acetyl-Neuraminidase Alpha-2,8-Sialyltransferase 6 (*ST8SIA6*), and the last SNP (rs4930390) is annotated to Chromosome 11 Open Reading Frame 80 (*C11orf80*). The minor alleles of the first 2 SNPs were associated with a higher risk of airflow obstruction and the minor allele of rs4930390 with a lower risk. The effect was significantly different between never- and ever-smokers for SNP rs4930390 according to both definitions and for rs7074210 in the FEV<sub>1</sub>/FVC < 70% analyses (see Additional file 1: Table S11). The three SNPs were not replicated in the SNP validation sample (see Table 3 and Additional file 1: Tables S9-S10).

**Gene expression in lung tissue**

The minor allele (G) of rs7519348 (overlapping SNP in never-smokers) was associated with higher gene expression of *NFYC* Antisense RNA 1 (*NFYC-AS1*) in lung tissue (Fig. 3). Summary statistics of the eQTL analysis for all overlapping SNPs at *p* < 10<sup>-4</sup> are provided in Additional file 2.

**Discussion**

We investigated the genetic overlap between GWASes using two airflow obstruction definitions in the same population (FEV<sub>1</sub>/FVC < 70 or < LLN). We expected a reasonable overlap in associated SNPs between the two definitions, since 96% of the never-smokers and 93% of the ever-smokers were classified the same way in the discovery sample LifeLines. Surprisingly, only a very small proportion (4% and 6%) of SNPs was overlapping at *p* < 10<sup>-4</sup> (see Fig. 1). Even with different significance thresholds the overlap was limited (26% and 29% at *p* < 0.05) (see Table 2). The same observation was made in the replication sample, the Vlagtwedde-Vlaardingen study. In this cohort, 94% and 90% of the never- and ever-smokers, respectively, were classified concordantly, but at *p* < 0.05 only 24% or 25% of the SNPs were overlapping. In addition, the effect estimates for the two airflow obstruction definitions correlated strongly in both cohorts but the *p*-values showed more variation and correlated only moderately resulting in different top-hits depending on the obstruction definition (see Fig. 2). Thus, the chosen strategy and definition of airflow obstruction had a substantial influence on the GWAS results. This implies that in a discovery-replication design with a predetermined selection *p*-value, different genetic variants would be followed-up depending on the definition used. In addition, there was no correlation between the *p*-values nor between the ORs of never- and ever-smokers in both cohorts. None of the selected SNPs overlapped between never- and



ever-smokers at  $p < 10^{-4}$ , and at  $p < 0.05$  the overlap was only 3% in LifeLines (discovery sample) and 2% in Vlagtwedde-Vlaardingen (replication sample, see Table 2). The current study therefore also highlights the importance of stratifying the analysis according to smoking status.

The difference between results from the two definitions might be explained by the fact that obstructive airway diseases are heterogeneous diseases with multiple phenotypes, symptoms and comorbidities. It might thus be beneficial for future GWA studies to focus more on specific COPD subtypes rather than on a broad definition

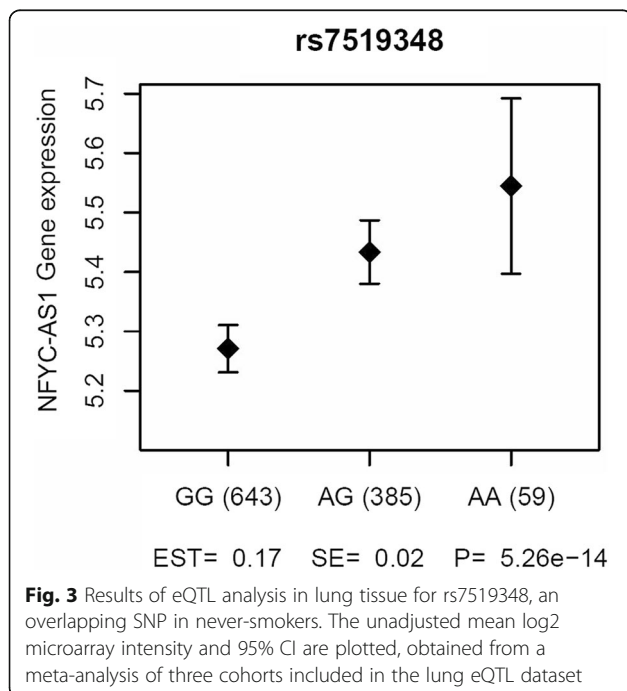
of airflow obstruction or COPD that can be caused by multiple underlying physiologic and genetic mechanisms. In previous GWA studies, in mainly smokers, on classical COPD phenotypes like emphysema and chronic bronchitis, the well-known general COPD genes (*HHIP*, *CHRNA* and *FAM13A*) were consistently identified [27–32]. Perhaps, to identify specific genetic pathways underlying specific COPD phenotypes we should not study the classical COPD phenotypes, but rather clinical COPD subtypes based on symptoms, comorbidities or pathology.

**Table 3** Results of the overlapping SNPs identified in both genome-wide association studies on FEV<sub>1</sub>/FVC < 70% and FEV<sub>1</sub>/FVC < LLN in never- and ever-smokers

SNP	Chr	A1	MAF	Gene	Test	Discovery analysis			SNP validation meta-analysis				Direction of effect in the independent cohorts <sup>a</sup>
						OR	SE	P	OR	SE	P	I <sup>2</sup>	
<i>Never-smokers</i>						<i>(n = 5070)</i>			<i>(n = 1966)</i>				
rs7519348	1	A	33%	NFYC (intronic)	<70%	1.36	0.07	4.92 × 10 <sup>-6</sup>	1.21	0.11	0.07	0.0	++++0
					<LLN	1.37	0.07	2.27 × 10 <sup>-5</sup>	1.40	0.16	0.03	0.0	+++++
rs6913003	6	T	4%	FABP7 (intronic)	<70%	1.90	0.13	8.99 × 10 <sup>-7</sup>	1.48	0.22	0.08	0.0	+++++
					<LLN	1.83	0.14	1.83 × 10 <sup>-5</sup>	1.72	0.31	0.08	0.0	+++++
<i>Ever-smokers</i>						<i>(n = 4855)</i>			<i>(n = 3134)</i>				
rs13118083	4	A	45%	HHIP (342 kb 5')	<70%	1.23	0.05	4.36 × 10 <sup>-5</sup>	0.97	0.07	0.60	17.0	+0 -- +
					<LLN	1.26	0.05	2.27 × 10 <sup>-5</sup>	1.05	0.08	0.57	46.7	++0 -- +
rs7074210	10	G	18%	ST8SIA6 (62 kb 5')	<70%	1.35	0.06	3.08 × 10 <sup>-6</sup>	1.05	0.08	0.57	64.7	+ - +++
					<LLN	1.33	0.07	3.23 × 10 <sup>-5</sup>	0.99	0.10	0.95	64.5	+ - + - +
rs4930390	11	G	24%	C11orf80 (intronic)	<70%	0.76	0.06	9.40 × 10 <sup>-6</sup>	1.02	0.07	0.74	0.0	- + 0 - 0
					<LLN	0.73	0.07	6.68 × 10 <sup>-6</sup>	1.01	0.09	0.91	10.1	- + + - -

SNPs were selected based on having a p-value < 10<sup>-4</sup> in both the discovery analyses on the fixed ratio of 70% and LLN. The logistic regression model of FEV<sub>1</sub>/FVC < 70% was adjusted for sex, age and height, the LLN model was not adjusted. Ever-smoking models were additionally adjusted for pack-years and current-smoking. Discovery sample = LifeLines cohort study, and SNP validation sample = Vlagtwedde-Vlaardingen and RS I to III. A1 = minor allele (effect allele), MAF = minor allele frequency, OR = Odds Ratio, SE = standard error and P = p-value, I<sup>2</sup> = heterogeneity measure  
<sup>a</sup> Order: LifeLines, Vlagtwedde-Vlaardingen, and Rotterdam Study I to III. + represents an OR > 1, - represents an OR < 1, and 0 represents an OR between 0.95 and 1.05 (no effect)

The *CHRNA5/3* and *HHIP* regions were overlapping between six previous GWA studies on airflow obstruction, using different airflow obstruction definitions and populations [11–16]. In the current study, two of the identified SNPs in ever-smokers were located in the *CHRNA5* and *HHIP* regions as well, pointing towards a



robust genetic association of these regions with airflow obstruction and COPD (see Additional file 1: Table S6). Likewise, most of previously identified regions associated with airflow obstruction or COPD were nominal significant ( $p < 0.05$ ) in the current study (see Additional file 1: Table S12). Out of the 22 loci identified by the study of Hobbs et al, SNPs in 18 loci were associated with at least one of the airflow definitions at a nominal significance (10 SNPs in never-smokers and 12 SNPs in ever smokers) [14]. In never-smokers, 6 of the 10 SNPs were significantly associated with both definitions and in ever-smokers 7 of the 12 SNPs were significantly associated with both definitions. Some SNPs were significant in both never- and ever-smokers (e.g. *HHIP*, *PID1* and *THSD4*), while others were either only significant in never-smokers (e.g. *FAM13A*, *DSP* and *RIN3*) or in ever-smokers (e.g. *CHRNA5*, *TET2* and *ADGRG6*). In addition, many of the loci previously associated with lung function outcomes (FEV<sub>1</sub>, FVC, and FEV<sub>1</sub>/FVC) were also nominal significant ( $p < 0.05$ ) in the current study (see Additional file 1: Table S13). Specifically, of the loci reported by Wain et al., 23 out of 28 loci for FEV<sub>1</sub>, 10 out of 17 loci for FVC and 38 out of 51 loci for FEV<sub>1</sub>/FVC were associated with at least one of the airflow definitions at a nominal significance [33]. Lastly, we also checked if the top overlapping SNPs were associated with lung function outcomes in our previous GWA studies on FEV<sub>1</sub>, FEV<sub>1</sub>/FVC and FEF<sub>25-75</sub> [34, 35]. A SNP annotated to *HHIP* was associated with FEV<sub>1</sub>/



FVC and  $FEF_{25-75}$  in both never- and ever-smokers (results were replicated) and the *CHRNA5/3* region was only associated with  $FEV_1/FVC$  in ever-smokers. The *NFYC* and *FABP7* regions were associated with  $FEV_1/FVC$  ( $p = 4.40 \times 10^{-4}$  and  $p = 1.87 \times 10^{-4}$ ) in never-smokers, and the *FABP7* SNP was also associated with  $FEF_{25-75}$  levels ( $p = 0.026$ ). Interestingly, the *NFYC* region was also overlapping between the current study and the study by Pillai et al. We identified multiple SNPs annotated to *NFYC*, whereas Pillai et al. identified a SNP (rs3767943) in the gene *KCNQ4*, which is located on the right side (3') of *NFYC* [15]. The *NFYC* region might therefore be an interesting region to further study the underlying mechanisms of its association with airflow obstruction.

A SNP in the intron of *NFYC* and a SNP in *FABP7* were the two overlapping SNPs between the airflow obstruction definitions at  $p < 10^{-4}$  in never-smokers and showed the same direction of effect in the five independent cohorts. The minor allele of the SNP in *NFYC* (rs7519348) was associated with a higher risk of airflow obstruction. This gene is a highly conserved transcription factor that is predicted by GeneGlobe to bind promoter regions of 218 genes (see Additional file 1: Table S14) including genes previously associated with lung related outcomes, like *ADORA2B*, *AKAP9*, *CDI63*, *ELMOD2*, *HLA-DPB1*, *ITPR2*, *KLF10* and *SERPINA6* [27, 36–42]. In more detail, *HLA-DPB1* is a known COPD gene related to disease severity, *SERPINA6* was associated with emphysema, a deletion in *ADORA2B* was shown to be associated with a decrease in lung fibrosis and pulmonary hypertension, and *ELMOD2* is a candidate gene for familial idiopathic pulmonary fibrosis [27, 36, 39, 40]. The identified SNP was not associated with expression levels of *NFYC* in lung tissue, but was an eQTL for a probeset annotated to *NFYC-AS1*. The function of this specific antisense-RNA, which are generally thought to have a regulatory role, is still unknown.

The minor allele of the SNP in *FABP7* (rs6913003) was also associated with a higher risk of airflow obstruction in never-smokers. This SNP was not associated with the expression of *FABP7* or other genes in lung tissue. *FABP7* is an intracellular lipid-binding protein, involved in long-chain fatty acids transport and cell proliferation [43]. It may be involved in abnormal pulmonary development, since lower expression of *FABP7* was found in patients with congenital cystic adenomatoid malformation [44]. In addition, higher expression of *FABP7* was seen in clear cell renal cell carcinoma and the authors suggested that the gene activates the *ERK* and *STAT3* signalling pathways [45]. *STAT3* was implicated to play a role in pulmonary inflammation and thus *FABP7* might indirectly be involved in airflow obstruction [46].

We were aware of the risk for spurious findings due to the low power of our study and thus we validated our top overlapping SNPs in 4 independent validation cohorts. We furthermore investigated the effect of low power on the overlap between the two definitions by increasing our dataset 2 and 4 times. We found that the percentage of overlap increases when the sample size increases, but still the number of SNPs that do not overlap remains high, i.e. 73.4% when the sample size increased 4-fold. So even when the study power is greatly increased, different SNPs will be found depending on the airflow obstruction definition tested. We also performed a simulation study by 10 times randomly allocating airflow obstruction cases and based on this simulation, we have to conclude that the differences and overlap we found could be chance findings, but that is why we validated the overlapping SNPs in 4 independent validation cohorts.

We only assessed a modest number of SNPs ( $n = 227,981$  SNPs) compared to previous large GWAS studies ( $n > 1$  million SNPs), since we only included genotyped SNPs to prevent any bias by imputation. The disadvantage of this approach is that we may have a lower genomic coverage. Another limitation of the current study is the use of pre-bronchodilator measurements to define airflow obstruction, which preferably should be based on post-bronchodilator measurements. Especially subjects with asthma could be misclassified as having airflow obstruction, but the results of the overlapping SNPs did not change in a sensitivity analysis excluding asthmatics or adjusting for asthma (see Additional file 1: Table S15). Moreover, only a low number of never-smoking subjects had an  $FEV_1/FVC < LLN$  in the three Rotterdam Study cohorts, but nevertheless results were replicated in these never-smokers. Finally, the “ $FEV_1/FVC < 70\%$ ” model was adjusted for sex, age and height, but the “ $FEV_1/FVC < LLN$ ” model was not adjusted for these variables, since they are included in the LLN calculation. If we do however adjust the “ $FEV_1/FVC < LLN$ ” model for these variables, the results do not change. The top SNPs are the same and the correlation between  $p$ -values for the LLN models adjusted and not adjusted is 0.98. In addition, the reported correlation in never-smokers between the two definitions was 0.48 for  $p$ -values and 0.78 for OR. If we use the LLN adjusted model the correlation is 0.48 and 0.79, respectively. This confirms that we used appropriate models to assess the genetic overlap between the two airflow definitions.

## Conclusions

The definition of airflow obstruction and the population under study may be important determinants of which SNPs are associated with airflow obstruction, and thus on which variants are selected for replication. It is

therefore important to use the same definition of airflow obstruction in future studies, especially in consortia. In addition, future studies should focus more on specific COPD subtypes and subgroups (e.g. based on smoking status), since there was no overlap in results between never- and ever-smokers, pointing towards possible different underlying mechanisms. Finally, our results suggest that the genes *FABP7* and *NFYC(-AS1)* could play a role in the pathogenesis of airflow obstruction in never-smokers.

## Additional files

**Additional file 1:** Supplementary methods, tables and figures. (DOCX 1790 kb)

**Additional file 2:** GWAS summary statistics. (XLSX 139809 kb)

## Abbreviations

COPD: Chronic Obstructive Pulmonary Disease; eQTL: Expression quantitative trait locus; *FABP7*: Fatty acid binding protein 7; FEV<sub>1</sub>/FVC: Ratio of forced expiratory volume in one second to forced vital capacity; GWAS: Genome-wide association study; MAF: Minor allele frequency; *NFYC*: Nuclear transcription factor Y subunit C; SNP: Single-nucleotide polymorphism

## Acknowledgments

We thank Rob Bieringa, Joost Keers, René Oostergo, Rosalie Visser, and Jan Schouten for their work related to data collection and validation in the LifeLines cohort study and the Vlagtwedde-Vlaardingen study. The authors would like to thank the staff at the Respiratory Health Network Tissue Bank of the FRQS for their valuable assistance with the lung eQTL dataset at Laval University. We are grateful to the study participants and all staff involved in the LifeLines cohort study, Vlagtwedde-Vlaardingen Study, Rotterdam Study and lung eQTL database. We would like to thank Anis Abuseiris, Karol Estrada, Dr. Tobias A. Knoch, and Rob de Graaf as well as their institutions Biophysical Genomics, Erasmus MC Rotterdam, The Netherlands, and especially the national German MediGRID and Services@MediGRID part of the German D-Grid, both funded by the German Bundesministerium fuer Forschung und Technology under grants #01 AK 803 A-H and # 01 IG 07015 G for access to their grid resources.

## Funding

This study is sponsored by grant number 4.1.13.007 of Lung Foundation (Longfonds), the Netherlands. DAvdP, KdJ and NA are supported by grant number 4.1.13.007 of Longfonds. LL is a Postdoctoral Fellow of the Research Foundation - Flanders (FWO). The LifeLines Biobank initiative has been made possible by funds from FES (Fonds Economische Structuurversterking), SNN (Samenwerkingsverband Noord Nederland) and REP (Ruimtelijke Economisch Programma). The Vlagtwedde-Vlaardingen cohort study was supported by the Ministry of Health and Environmental Hygiene of the Netherlands and the Netherlands Asthma Fund (grant 187) and the genetic data of the cohort were funded by the Netherlands Asthma Fund (grant no. 3.2.02.51), the Stichting Astma Bestrijding, BBMRI-NL (Complementation project) and the European Respiratory Society COPD research award 2011 to H.M. Boezen. The Rotterdam Study was supported by the Erasmus MC and Erasmus University Rotterdam; the Netherlands Organisation for Scientific Research (NWO); the Netherlands Organisation for Health Research and Development (ZonMw); the Research Institute for Diseases in the Elderly (RIDE); the Netherlands Genomics Initiative (NGI); the Ministry of Education, Culture and Science; the Ministry of Health, Welfare and Sports; the European Commission (DG XII); and the Municipality of Rotterdam. Genotyping and gene expression for the lung eQTL study was funded by Merck & Co. Inc. The lung eQTL study at Laval University was supported by the Fondation de l'Institut universitaire de cardiologie et de pneumologie de Québec, the Respiratory Health Network of the FRQS, the Canadian Institutes of Health Research (MOP - 123369). Y.B. holds a Canada Research Chair in Genomics of Heart and Lung Diseases. The sponsors

of this study played no role in the design of the study, data collection, analysis, interpretation or in the writing and submission of the manuscript.

## Availability of data and materials

Summary statistics of all analyses are available in Additional file 2. The eQTL lung tissue datasets analysed during the current study are available at the Gene Expression Omnibus (GEO) repository, GEO accession numbers GSE23546 and GPL10379. LifeLines data is available (at costs) to all scientists. Scientists can apply for access to Lifelines data and samples by submitting a research proposal to the LifeLines biobank ([www.lifelines.net](http://www.lifelines.net)).

## Authors' contributions

DAvdP participated in the study design, analysis and interpretation of the data, and drafting of the manuscript, tables and figures. HMB, DSP, CCvD and CMVD obtained funding. KdJ, HMB, DSP, JMV, CCvD, CMVD, IN and NA determined the study design, participated in the analysis and interpretation of data, and critically supervised writing of the manuscript. LL and GGB participated in collecting and analysing the Rotterdam Study data. AF, YB, CAB, DSP, KH, and PDP participated in setting up the lung tissue database and analyses. All authors approved the final version of the manuscript.

## Ethics approval and consent to participate

Written informed consent was provided by all subjects of all included cohorts. The Lifelines Cohort Study and the Vlagtwedde-Vlaardingen study were approved by the Medical Ethics Committee of the University Medical Center Groningen, Groningen, the Netherlands. The Rotterdam Study was approved by the medical ethics committee of Erasmus University. The eQTL lung tissue database was approved by the ethics committees of the Institut universitaire de cardiologie et de pneumologie de Québec (Laval) and the UBC-Providence Health Care Research Institute Ethics Board (UBC). The study protocol was consistent with the Research Code of the University Medical Center Groningen and Dutch national ethical and professional guidelines.

## Consent for publication

Not applicable.

## Competing interests

LL reports personal fees from Boehringer Ingelheim GmbH, non-financial support from Novartis, grants from AstraZeneca, grants and non-financial support from European Respiratory Society and Belgian Respiratory Society, outside the submitted work. The University of Groningen has received money for Professor Postma (DSP) regarding a grant for research from Astra Zeneca, Chiesi, Genentec, GSK and Roche. Fees for consultancies were given to the University of Groningen by Astra Zeneca, Boehringer Ingelheim, Chiesi, GSK, Takeda and TEVA. All other authors declare no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>Department of Epidemiology, University of Groningen, University Medical Center Groningen, Hanzeplein 1, 9700, RB, Groningen, The Netherlands.

<sup>2</sup>Groningen Research Institute for Asthma and COPD (GRIAC), University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

<sup>3</sup>Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands.

<sup>4</sup>Department of Respiratory Medicine, Ghent University Hospital, Ghent, Belgium.

<sup>5</sup>Department of Pathology and Medical Biology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

<sup>6</sup>Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

<sup>7</sup>Department of Respiratory Medicine, Erasmus Medical Center, Rotterdam, The Netherlands.

<sup>8</sup>Department of Molecular Medicine, Institut universitaire de cardiologie et de pneumologie de Québec, Laval University, Québec, Canada.

<sup>9</sup>Merck Research Laboratories, Boston, MA, USA.

<sup>10</sup>Department of Medicine, Center for Heart Lung Innovation and Institute for Heart and Lung Health, University of British Columbia, St. Paul's Hospital, Vancouver, Canada.

<sup>11</sup>Department of Pulmonary Diseases, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.

Received: 13 June 2018 Accepted: 11 February 2019  
 Published online: 07 March 2019

References

1. World Health Organisation (WHO). The top 10 causes of death, Fact sheet N°310. <http://www.who.int/mediacentre/factsheets/fs310/en/>.
2. Postma DS, Kerkhof M, Boezen HM, Koppelman GH. Asthma and chronic obstructive pulmonary disease: common genes, common environments? *Am J Respir Crit Care Med*. 2011;183:1588–94. <https://doi.org/10.1164/rccm.201011-1796PP>.
3. Global initiative for chronic Obstructive Lung Disease (GOLD). Global Strategy for the Diagnosis, Management and Prevention of COPD 2017. 2017. <http://www.goldcopd.org/>.
4. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. *Eur Respir J*. 2005;26:948–68.
5. Mohamed Hoesein FAA, Zanen P, Lammers J-WJ. Lower limit of normal or FEV1/FVC <0.70 in diagnosing COPD: An evidence-based review. *Respir Med*. 2011;105:–907, 15. <https://doi.org/10.1016/j.rmed.2011.01.008>.
6. Medbo A, Melbye H. Lung function testing in the elderly—can we still use FEV1/FVC<70% as a criterion of COPD? *Respir Med*. 2007;101:1097–105.
7. Hardie JA, Buist AS, Vollmer WM, Ellingsen I, Bakke PS, Morkve O. Risk of over-diagnosis of COPD in asymptomatic elderly never-smokers. *Eur Respir J*. 2002;20:1117–22.
8. Roberts SD, Farber MO, Knox KS, Phillips GS, Bhatt NY, Mastronarde JG, et al. FEV1/FVC ratio of 70% misclassifies patients with obstruction at the extremes of age. *Chest*. 2006;130:200–6.
9. Sorino C, D’Amato M, Steinhilber G, Patella V, Corsico AG. Spirometric criteria to diagnose airway obstruction in the elderly: fixed ratio vs lower limit of normal. *Minerva Med*. 2014;105(6 Suppl 3):15–21.
10. Ramsey SD. Suboptimal medical therapy in COPD: exploring the causes and consequences. *Chest*. 2000;117(2 Suppl):33S–7S.
11. Cho MH, Boutaoui N, Klanderman BJ, Sylvia JS, Ziniti JP, Hersh CP, et al. Variants in FAM13A are associated with chronic obstructive pulmonary disease. *Nat Genet*. 2010;42:200–2. <https://doi.org/10.1038/ng.535>.
12. Cho MH, Castaldi PJ, Wan ES, Siedlinski M, Hersh CP, Demeo DL, et al. A genome-wide association study of COPD identifies a susceptibility locus on chromosome 19q13. *Hum Mol Genet*. 2012;21:947–57. <https://doi.org/10.1093/hmg/ddr524>.
13. Cho MH, McDonald ML, Zhou X, Mattheisen M, Castaldi PJ, Hersh CP, et al. Risk loci for chronic obstructive pulmonary disease: a genome-wide association study and meta-analysis. *Lancet Respiratory Med*. 2014;2:214–25. [https://doi.org/10.1016/S2213-2600\(14\)70002-5](https://doi.org/10.1016/S2213-2600(14)70002-5).
14. Hobbs BD, de Jong K, Lamontagne M, Bosse Y, Shrine N, Artigas MS, et al. Genetic loci associated with chronic obstructive pulmonary disease overlap with loci for lung function and pulmonary fibrosis. *Nat Genet*. 2017;49:426–32. <https://doi.org/10.1038/ng.3752>.
15. Pillai SG, Ge D, Zhu G, Kong X, Shianna KV, Need AC, et al. A genome-wide association study in chronic obstructive pulmonary disease (COPD): identification of two major susceptibility loci. *PLoS Genet*. 2009;5:e1000421. <https://doi.org/10.1371/journal.pgen.1000421>.
16. Wilk JB, Shrine NR, Loehr LR, Zhao JH, Manichaikul A, Lopez LM, et al. Genome-wide association studies identify CHRNA5/3 and HTR4 in the development of airflow obstruction. *Am J Respir Crit Care Med*. 2012;186:622–32. <https://doi.org/10.1164/rccm.201202-0366OC>.
17. Stolk RP, Rosmalen JG, Postma DS, de Boer RA, Navis G, Slaets JP, et al. Universal risk factors for multifactorial diseases: LifeLines: a three-generation population-based study. *Eur J Epidemiol*. 2008;23:67–74. <https://doi.org/10.1007/s10654-007-9204-4>.
18. van der Lende R Gezonheidsorganisatie T.N.O. R te G. Epidemiology of chronic non-specific lung disease (chronic Bronchitis). A critical analysis of three field surveys of CNSLD carried out in the Netherlands. Van Gorcum; 1969.
19. Gosman MM, Boezen HM, van Diemen CC, Snoeck-Stroband JB, Lapperre TS, Hiemstra PS, et al. A disintegrin and metalloprotease 33 and chronic obstructive pulmonary disease pathophysiology. *Thorax*. 2007;62:242–7.
20. Hofman A, Brusselle GG, Darwishi Murad S, van Duijn CM, Franco OH, Goedegebure A, et al. The Rotterdam Study: 2016 objectives and design update. *Eur J Epidemiol*. 2015;30:661–708. <https://doi.org/10.1007/s10654-015-0082-x>.
21. Quanjer PH, Stanojevic S, Cole TJ, Baur X, Hall GL, Culver BH, et al. Multi-ethnic reference values for spirometry for the 3–95-yr age range: the global lung function 2012 equations. *Eur Respir J*. 2012;40:1324–43. <https://doi.org/10.1183/09031936.00080312>.
22. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26:319–338.
23. Khankhanian P, Din L, Caillier SJ, Gourraud PA, Baranzini SE. SNP imputation bias reduces effect size determination. *Front Genet*. 2015. <https://doi.org/10.3389/fgene.2015.00030>.
24. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
25. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*. 2010;26:2190–1. <https://doi.org/10.1093/bioinformatics/btq340>.
26. Hao K, Bosse Y, Nickle DC, Pare PD, Postma DS, Laviolette M, et al. Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet*. 2012;8:e1003029. <https://doi.org/10.1371/journal.pgen.1003029>.
27. Kong X, Cho MH, Anderson W, Coxson HO, Muller N, Washko G, et al. Genome-wide association study identifies BICD1 as a susceptibility gene for emphysema. *Am J Respir Crit Care Med*. 2011;183:43–9. <https://doi.org/10.1164/rccm.201004-0541OC>.
28. Manichaikul A, Hoffman EA, Smolonska J, Gao W, Cho MH, Baumhauer H, et al. Genome-wide study of percent emphysema on computed tomography in the general population. The Multi-Ethnic Study of Atherosclerosis Lung/SNP Health Association Resource Study. *Am J Respir Crit Care Med*. 2014;189:408–18. <https://doi.org/10.1164/rccm.201306-1061OC>.
29. Castaldi PJ, Cho MH, San Jose Estepar R, McDonald ML, Laird N, Beaty TH, et al. Genome-wide association identifies regulatory Loci associated with distinct local histogram emphysema patterns. *Am J Respir Crit Care Med*. 2014;190:399–409. <https://doi.org/10.1164/rccm.201403-0569OC>.
30. Cho MH, Castaldi PJ, Hersh CP, Hobbs BD, Barr RG, Tal-Singer R, et al. A Genome-Wide Association Study of Emphysema and Airway Quantitative Imaging Phenotypes. *Am J Respir Crit Care Med*. 2015;192:559–69. <https://doi.org/10.1164/rccm.201501-0148OC>.
31. Boueiz A, Lutz SM, Cho MH, Hersh CP, Bowler RP, Washko GR, et al. Genome-Wide Association Study of the Genetic Determinants of Emphysema Distribution. *Am J Respir Crit Care Med*. 2017;195:757–71. <https://doi.org/10.1164/rccm.201605-0997OC>.
32. Lee JH, Cho MH, Hersh CP, McDonald ML, Crapo JD, Bakke PS, et al. Genetic susceptibility for chronic bronchitis in chronic obstructive pulmonary disease. *Respir Res*. 2014;15:112–3. <https://doi.org/10.1186/s12931-014-0113-2>.
33. Wain LV, Shrine N, Artigas MS, Erzurumluoglu AM, Noyvert B, Bossini-Castillo L, et al. Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat Genet*. 2017;49:416–25. <https://doi.org/10.1038/ng.3787>.
34. van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, et al. Genome-wide association study on the FEV1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J Allergy Clin Immunol*. 2017;139:533–40. <https://doi.org/10.1016/j.jaci.2016.06.062>.
35. van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, et al. The well-known gene HHIP and novel gene MECR are implicated in small airway obstruction. *Am J Respir Crit Care Med*. 2016;194. <https://doi.org/10.1164/rccm.201604-0843LE>.
36. Karmouty-Quintana H, Philip K, Acero LF, Chen NY, Weng T, Molina JG, et al. Deletion of ADORA2B from myeloid cells dampens lung fibrosis and pulmonary hypertension. *FASEB J*. 2015;29:50–60. <https://doi.org/10.1096/fj.14-260182>.
37. Oldenburger A, Poppinga WJ, Kos F, de Bruin HG, Rijks WF, Heijink IH, et al. A-kinase anchoring proteins contribute to loss of E-cadherin and bronchial epithelial barrier by cigarette smoke. *Am J Physiol Physiol*. 2014;306:C585–97. <https://doi.org/10.1152/ajpcell.00183.2013>.
38. Kaku Y, Imaoka H, Morimatsu Y, Komohara Y, Ohnishi K, Oda H, et al. Overexpression of CD163, CD204 and CD206 on alveolar macrophages in the lungs of patients with severe chronic obstructive pulmonary disease. *PLoS One*. 2014;9:e87400. <https://doi.org/10.1371/journal.pone.0087400>.
39. Hodgson U, Pulkkinen V, Dixon M, Peyrard-Janvid M, Rehn M, Lahermo P, et al. ELMOD2 is a candidate gene for familial idiopathic pulmonary fibrosis. *Am J Hum Genet*. 2006;79:149–54. <https://doi.org/10.1086/504639>.
40. Savarimuthu Francis SM, Larsen JE, Pavey SJ, Duhig EE, Clarke BE, Bowman RV, et al. Genes and gene ontologies common to airflow obstruction and emphysema in the lungs of patients with COPD. *PLoS One*. 2011;6:e17442. <https://doi.org/10.1371/journal.pone.0017442>.

41. Wilker EH, Alexeeff SE, Poon A, Litonjua AA, Sparrow D, Vokonas PS, et al. Candidate genes for respiratory disease associated with markers of inflammation and endothelial dysfunction in elderly men. *Atherosclerosis*. 2009;206:480–5. <https://doi.org/10.1016/j.atherosclerosis.2009.03.004>.
42. Koczulla AR, Jonigk D, Wolf T, Herr C, Noeske S, Klepetko W, et al. Kruppel-like zinc finger proteins in end-stage COPD lungs with and without severe alpha1-antitrypsin deficiency. *Orphanet J Rare Dis*. 2012;7:29. <https://doi.org/10.1186/1750-1172-7-29>.
43. Haunerland NH, Spener F. Fatty acid-binding proteins—insights from genetic manipulations. *Prog Lipid Res*. 2004;43:328–49. <https://doi.org/10.1016/j.plipres.2004.05.001>.
44. Wagner AJ, Stumbaugh A, Tigue Z, Edmondson J, Paquet AC, Farmer DL, et al. Genetic analysis of congenital cystic adenomatoid malformation reveals a novel pulmonary gene: fatty acid binding protein-7 (brain type). *Pediatr Res*. 2008;64:11–6. <https://doi.org/10.1203/PDR.0b013e318174eff8>.
45. Zhou J, Deng Z, Chen Y, Gao Y, Wu D, Zhu G, et al. Overexpression of FABP7 promotes cell growth and predicts poor prognosis of clear cell renal cell carcinoma. *Urol Oncol*. 2015;33:113.e9–113.17. <https://doi.org/10.1016/j.urolonc.2014.08.001>.
46. Ruwanpura SM, McLeod L, Miller A, Jones J, Vlahos R, Ramm G, et al. Deregulated Stat3 signaling dissociates pulmonary inflammation from emphysema in gp130 mutant mice. *Am J Physiol Cell Mol Physiol*. 2012;302:L627–39. <https://doi.org/10.1152/ajplung.00285.2011>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

